

Durham Research Online

Deposited in DRO:

03 October 2016

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Katramados, I. and Breckon, T.P. (2016) 'Dense gradient-based features (DeGraF) for computationally efficient and invariant feature extraction in real-time applications.', in 2016 IEEE International Conference on Image Processing (ICIP), September 25-28, 2016, Phoenix, Arizona, USA ; proceedings. Piscataway, NJ: IEEE, pp. 300-304.

Further information on publisher's website:

<http://dx.doi.org/10.1109/ICIP.2016.7532367>

Publisher's copyright statement:

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

DENSE GRADIENT-BASED FEATURES (DEGRAF) FOR COMPUTATIONALLY EFFICIENT AND INVARIANT FEATURE EXTRACTION IN REAL-TIME APPLICATIONS

Ioannis Katramados*, Toby P. Breckon†

*{Cranfield University | Cosmonio Ltd}, Bedfordshire, UK †Durham University, Durham, UK

ABSTRACT

We propose a computationally efficient approach for the extraction of dense gradient-based features based on the use of localized intensity-weighted centroids within the image. Whilst prior work concentrates on sparse feature derivations or computationally expensive dense scene sensing, we show that Dense Gradient-based Features (DeGraF) can be derived based on initial multi-scale division of Gaussian preprocessing, weighted centroid gradient calculation and either local saliency (DeGraF- α) or signal-to-noise inspired (DeGraF- β) final stage filtering. We present two variants (DeGraF- α / DeGraF- β) of which the signal-to-noise based approach is shown to perform admirably against the state of the art in terms of feature density, computational efficiency and feature stability. Our approach is evaluated under a range of environmental conditions typical of automotive sensing applications with strong feature density requirements.

Index Terms— dense features, feature invariance, feature points, intensity weighted centroids, automotive vision

1. INTRODUCTION

The derivation of stable and robust point features has long been the focus of significant research effort underpinning the requirements of object recognition, tracking and depth estimation (structure from motion), sparse stereo vision and generalized scene classification. Early promising work [1, 2, 3], followed by the seminal Scale Invariant Feature Transform (SIFT) [4], led to industrious efforts to optimize sparse feature stability against the potential for real-time performance:- FAST [5], SURF [6], Hessian-affine [7, 8], MSER [9], Kadir-Brady [10], EBR [11], IBR [11], CenSurE [12], AGAST [13], ASIFT [14], BRIEF [15], ORB [16], FREAK [17], BRISK [18].

Against the backdrop of a standardized set of feature stability metrics proposed from [8], the wider applicability of newer sparse feature point contenders to the field has narrowed to variants that either expunged strong orientational invariance (e.g. BRIEF [15]), concentrated on computationally robust sparse matching (FREAK [17], BRISK [18]) or instead improved stability by introducing additional computation (e.g. ASIFT, [14]). Although with their merits, the rise of autonomous platform driven applications [19] and the rise of sensing tasks in the space of real-time scene classification [20], visual odometry [21], monocular depth estimation [22], scene flow [23] and alike instead gave rise to the need for dense features (Figure 1). The immediate response of simply



Fig. 1: Dense DeGraF- β features (right) from vehicle sensing (left) using dense feature grids (e.g. dense SIFT, [24]) found itself falling foul of the recent trend in feature point optimization - *improved feature stability at the expense of computational cost Vs. reduced computational cost at the expense of feature stability*. Such applications found themselves instead requiring real-time high density features that were in themselves stable to the narrower subset of metric conditions typical of the automotive-type application genre (e.g. lesser camera rotation, high image noise, extreme illumination changes).

Whilst contemporary work aimed to address this issue via dense scene mapping [23, 22] (computationally enabled by GPU) or a move to scene understanding via 3D stereo sensing [25, 26], here we specifically consider applicability to future low-carbon (possibly electric), long-duration autonomous vehicles where Size, Weight and (computational) Power (SWaP) requirements will heavily govern the design process. Using a wide range of metrics proposed by Mikolajczyk et al. [8], our evaluation (Section 3) shows that there is a gap for a computationally efficient feature detection approach that produces high density invariant features (unlike the FREAK / BRISK contenders [18, 17]). To these ends, we present a computationally lightweight feature detection approach that is shown to be highly invariant (stable) under conditions typical of both automotive and wider autonomous sensing.

2. DENSE GRADIENT-BASED FEATURES (DEGRAF)

A novel feature detection approach, denoted as DeGraF (Dense Gradient Features), is presented based on the calculation of gradients using intensity-weighted centroids [27].

2.1. Gradients from Centroids (GraCe)

Firstly, we identify a novel method for calculating dense image gradients using intensity-weighted centroid as an altern-

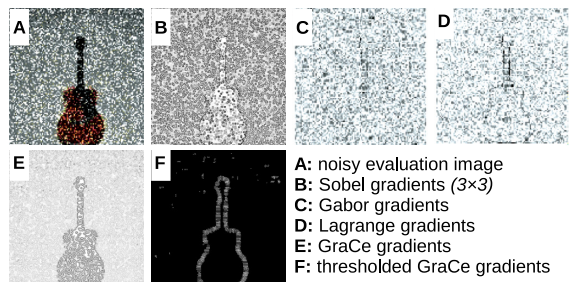


Fig. 2: Comparison of gradient calculation operators: (A-D [28])

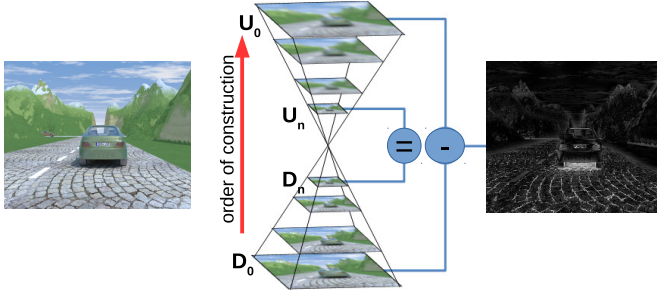


Fig. 3: Difference of Gaussian calculation using a di-pyramid.

ative to commonplace Sobel operator used in contemporary work [29, 30]. In this context, for any $w \times h$ dimension image region, I , with spatial centroid $I(x_c, y_c)$, we can define two symmetrical centroids, C_{pos} and C_{neg} , that define a gradient flow vector $\vec{C_{neg}C_{pos}}$ as the spatially weighted average pixel value via Eqn. 1:

$$C_{pos}(x_{pos}, y_{pos}) = C_{pos} \left(\frac{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} i I(i, j)}{S_{pos}}, \frac{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} j I(i, j)}{S_{pos}} \right) \quad (1)$$

$$C_{neg}(x_{neg}, y_{neg}) =$$

$$C_{neg} \left(\frac{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} i(1+m-I(i, j))}{S_{neg}}, \frac{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} j(1+m-I(i, j))}{S_{neg}} \right)$$

where $S_{pos} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} I(i, j)$, $S_{neg} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} (1+m-I(i, j))$ and $m = \max_{(i,j)} I(i, j)$. Here the negative centroid is defined as the weighted average of the inverted pixel values (Eqn. 1) with pixel values normalized ($1 \rightarrow 256$) to avoid division by zero. To avoid instability within the resulting gradient vector, $\vec{C_{neg}C_{pos}}$, in noisy image regions we ensure $S_{pos} > S_{neg}$. If not, we subsequently redefine both symmetrical centroids, C_{pos} and C_{neg} , such that $C'_{pos} = C_{neg}()$ (from Eqn. 1) and C'_{neg} is the symmetric point about the spatial centroid $I(x_c, y_c)$ (Eqn. 2) such as to ensure $S_{pos} > S_{neg}$:

$$C'_{neg}(x_{neg}, y_{neg}) = (2x_c - x_{pos}, 2y_c - y_{pos}) \quad (2)$$

This correction dramatically increases the accuracy of both the resulting orientation, $\phi = \tan^{-1}(dy, dx)$, and magnitude, $r = \sqrt{dx^2 + dy^2}$, as calculated based on $\frac{dx}{dy} = \frac{x_{pos} - x_{neg}}{y_{pos} - y_{neg}}$ in 2D image space, from resulting gradient vector $\vec{C_{pos}C_{neg}}$ (or $\vec{C'_{pos}C'_{neg}}$) under noise. As Figure 2 illustrates, this approach has several advantages over conventional Sobel, Gabor and Lagrange gradients [28] in that it is invariant to region size, offers sub-pixel accuracy and has significantly higher noise resistance. Computationally it remains more efficient than Gabor and Lagrange approaches and only marginally more complex than Sobel mask convolution.

2.2. From Gradients to Features

Dense gradient-based features (DeGraF) are extracted, based on our GraCe gradient map, using a three step process.

Difference of Gaussian (DoG) is used as an (optional) illumination invariance image representation (akin to [31]). This is derived using an *Inverted Gaussian Di-pyramid* concept (Fig. 3). A di-pyramid conceptually comprises two pyramids symmetrically placed base-to-base, whereas an inverted di-pyramid comprises two pyramids symmetrically placed peak-to-peak (Fig. 3). The original image is used to perform bottom-up construction of a classical (down-sampling) Gaussian pyramid, D , constructed with n levels (Fig. 3, lower pyramid). Subsequently, the peak of this pyramid, the n^{th} level D_n , is used to perform top-down construction of a second (up-sampling) Gaussian pyramid, U with base U_0 , from a starting point of $U_n = D_n$ (Fig. 3, upper inverted pyramid). Within this inverted di-pyramid arrangement, calculating the absolute difference between the two pyramid bases gives a DoG image, $I_{DoG} = |U_0 - D_0|$, as an intermediate saliency based scene representation.

Gradient matrix calculation is performed using a $w_G \times h_G$ dimension grid G overlain onto the DoG (or original) image. A GraCe gradient vector, $\vec{C_{neg}C_{pos}}$, is calculated for each grid cell C , with cell dimension $w_C \times h_C$. Cells overlap by δ_x pixels horizontally and δ_y pixels vertically such that the dimensions of the resulting gradient matrix G are as follows:

$$w_G = \left\lfloor \frac{w - \delta_x}{w_C - \delta_x} \right\rfloor \quad h_G = \left\lfloor \frac{h - \delta_y}{h_C - \delta_y} \right\rfloor \quad (3)$$

where $0 \leq \delta_x \leq w_c - 1$ and $0 \leq \delta_y \leq h_c - 1$. Choice of gradient matrix parameters depend on required computational performance, feature density and noise resistance. Figure 4 illustrates varying gradient vectors with different cell size and overlap.

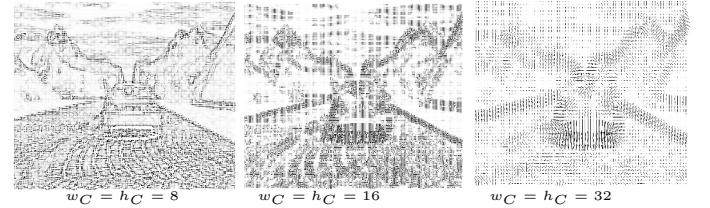


Fig. 4: Gradient matrix calculation ($\delta_x = \delta_y = 4$).

Feature extraction is the subsequent process of down-selecting those gradient vectors that are suitable for sequential frame-to-frame tracking. Global feature extraction using α -gradients uses an approach similar to CenSurE [12] and SIFT [4], where local minima and extrema are identified but here within the gradient matrix derived by GraCe (Section 2.1) An α -gradient is defined as being locally salient with either a consistently stronger or weaker gradient magnitude in comparison to its immediate spatial neighbours in G . Local feature extraction using β -gradients is an alternative to extracting a salient subset of gradients and instead looks to identify a subset of gradients where the Signal-to-Noise (SNR) ratio is low ($SNR = 20 \log_{10} \frac{\text{signal}}{\text{RMS noise}} \approx \frac{\mu_{sig}}{\sigma_{bg}}$). By returning to Eqn. 1 we can see that our positive and negative centroid, C_{pos} and C_{neg} , give rise to both a measure of the maximal signal, $S_{pos} \approx \mu_{sig}$, and minimal signal, $S_{neg} \approx \sigma_{bg}$, when

the gradient matrix calculation is itself performed on the DoG image (high $\approx \mu_{sig}$; low $\approx \sigma_{bg}$, see Fig. 3). By extension, with the DoG image itself being akin to a second-order saliency derivative of the original input, this also holds for input image itself. To these ends, we can define gradient quality metrics such that gradient vectors, $\vec{C_{neg}C_{pos}}$, with corresponding magnitude, r , less than a given threshold, $r < \tau_r$, are discarded as are those with a low centroid ratio, R , measuring signal-to-noise ratio for each of the centroids (Eqn. 4).

$$R = \min \left(\frac{S_{pos}}{S_{neg}}, \frac{S_{neg}}{S_{pos}} \right) \quad (4)$$

Using either feature extraction approach, the proposed DeGraF approach derives stable features from a gradient matrix using intensity-weighted centroids. This enforces uniformity of feature distribution across the image with density configurable via the parameters of G (see Figure 4). This approach is primarily designed for deriving noise resistant and high density features within environments where conventional feature extraction approaches fail. GraCe (Section 2.1) introduces a novel way of handling noise by separately estimating the signal-to-noise ratio for each gradient. In this context, a gradient is defined as the combination of a positive and a negative centroid. Since the weakest of the two centroids is more vulnerable to noise, only the dominant centroid is considered. The weakest centroid is then redefined as the anti-symmetric opposite to the positive centroid (Section 2.1). This approach of calculating gradients is the key contributing factor to the subsequent performance of DeGraF features. Furthermore, extracting DeGraF features from a DoG image adds additional illumination invariance with the possibility of extension across multiple DoG pyramids levels for higher feature density and scale-invariance. Crucially, dense DeGraF features retain low and tunable computational complexity facilitating real-time computation. Furthermore, the grid-wise computational approach readily facilitates parallelization (multi-core, GPU, FPGA and alike).

3. EVALUATION

Our DeGraF feature detection is compared to a number of state-of-the-art approaches:- Adaptive and Generic Corner Detection Based on the Accelerated Segment Test (**AGAST**) [13], Centre Surround Extremas for Real-time Feature Detection and Matching (**CenSurE**) [12], Features from Accelerated Segment Test (**FAST**) [5], Good Features To Track (**GFTT**) [3], Scale Invariant Feature Transform (**SIFT**) [4], Speeded Up Robust Features (**SURF**) [6], Oriented Fast and Rotated BRIEF (**ORB**) [16] (superseding [15]) and Maximally Stable Extremal Regions (**MSER**) [9]. These are compared to the two proposed variants **DeGraF- α** and **DeGraF- β** .

Quantitative evaluation is based on six statistical measures:- feature density (as % of image resolution, Table 1), tracking accuracy (as % of total features tracked successfully between

| | density (%) | execution (ms.) | frame-rate (fps) |
|---|-------------|-----------------|------------------|
| AGAST [13] | 4.35 | 33 | 30 |
| (b = 1, with non-maximal suppression) | | | |
| CenSurE [12] | 0.44 | 27 | 37 |
| (max. = 5, res. thres. = 0, projected = 10, binarised = 5, sup. non-max. = 2) | | | |
| DeGraF- α | 4.47 | 35 | 28 |
| (w _G = h _G =2, $\delta_x = \delta_y = 1$, using DoG) | | | |
| DeGraF- β | 6.07 | 24 | 41 |
| (w _G = h _G =3, $\delta_x = \delta_y = 3$, $\tau_r = 0.015$, using DoG) | | | |
| FAST [5] | 3.41 | 20 | 50 |
| (threshold = 0, with non-maximal suppression) | | | |
| GFTT [3] | 3.18 | 43 | 23 |
| (max. corners = 0, quality = 0.001, min. dist. = 1.0, b = 3, k = 0.04) | | | |
| SIFT [4] | 3.84 | 109 | 9 |
| (octaves = 3, contrast = 0.015, edge = 10, $\sigma = 0.7$) | | | |
| SURF [6] | 0.97 | 39 | 25 |
| (hessian = 0, octaves = 4, octave layers = 2) | | | |
| ORB [16] | 2.96 | 37 | 27 |
| (scale = 1.2, levels = 8, patch size = 1, with Harris response) | | | |
| MSER [9] | 0.19 | 293 | 3 |
| ($\delta = 5$, min/max: 60/1000, var.: 0.25, min. div.: 0.2, max evol.: 200, area: 1.01, min marg.: 0.003, blur: 5) | | | |

Table 1: Feature density, execution time and frame-rate.

subsequent frames, Table 2), feature matching repeatability under variable illumination (% error in features illumination variant matching, Table 2), feature matching repeatability under variable rotation (% error in rotational features matching, Table 3), feature matching repeatability under additive Gaussian noise (% error in features matching introduced, Table 2) and execution time (mean per frame (milliseconds) / frame-rate (frames per second, fps): single-core 2.4 Ghz CPU on 640×480 resolution images, Table 1). These are a variant on the criteria of Mikolajczyk et al. [11], adapted to the requirements of feature sparse environments typical of automotive visual sensing applications. For example, an on-vehicle camera is expected to vibrate or rotate slightly but this will not result in significant affine transformations as in the general case of [11]. Our statistical evaluation is performed on the Enpeda EISATS dataset [32, 23] facilitating accurate ground truth data independent of scene noise.

Table 1 shows comparative feature density and execution time where we can see DeGraF- β produces the highest density (to be expected since the entire gradient matrix is used to produce one feature per positive centroid) and second lowest execution time. The second highest feature density is shown by DeGraF- α which is comparable to AGAST in terms of density and execution time. Other feature detectors produced lower feature density responses with FAST, DeGraF- β and CenSurE providing the lowest execution times.

Feature tracking accuracy is tested within an automotive context with the use of artificially added camera vibration, of amplitude $\pm v$ pixels, in the vertical direction. Pyramidal Lucas-Kanade tracking [33] is then used to track features between frames with mean feature tracking error defined as:

| | vibration amplitude $\pm v$ pixels | | | | | | % increase in illumination, Δi | | | | image noise, $n\%$ of pixel effected | | | |
|----------------------------------|------------------------------------|-------------|-------------|--------------|--------------|--------------|--|---------------|---------------|---------------|--------------------------------------|---------------|---------------|---------------|
| | ± 1 | ± 2 | ± 4 | ± 8 | ± 16 | ± 32 | 25% | 50% | 75% | 100% | 5% | 10% | 15% | 20% |
| AGAST [13] | 3.64 | 7.44 | 9.39 | 18.44 | 32.83 | 36.03 | 36.90% | 41.58% | 46.73% | 52.11% | 70.24% | 78.03% | 79.92% | 82.13% |
| CenSurE [12] | 5.02 | 6.82 | 6.74 | 16.08 | 28.08 | 30.80 | 33.62% | 43.41% | 51.90% | 59.21% | 74.60% | 86.22% | 90.28% | 92.39% |
| DeGraF- α | 0.07 | 0.17 | 3.56 | 14.39 | 26.13 | 36.48 | 17.68% | 21.25% | 25.96% | 30.79% | 59.22% | 74.27% | 79.42% | 83.14% |
| DeGraF-β | 0.36 | 0.47 | 2.25 | 10.34 | 24.62 | 34.91 | 8.11% | 17.93% | 25.44% | 30.49% | 3.68% | 7.86% | 10.72% | 13.84% |
| FAST [5] | 3.90 | 7.80 | 10.59 | 18.96 | 34.27 | 36.88 | 38.60% | 43.96% | 49.25% | 55.23% | 73.62% | 80.19% | 83.13% | 84.00% |
| GFTT [3] | 4.00 | 8.54 | 9.71 | 17.63 | 32.73 | 37.17 | 28.51% | 37.71% | 45.49% | 54.56% | 75.82% | 82.64% | 84.99% | 86.36% |
| SIFT [4] | 1.80 | 4.39 | 6.76 | 16.95 | 35.34 | 35.44 | 28.62% | 37.23% | 46.40% | 55.11% | 68.24% | 80.61% | 83.60% | 85.05% |
| SURF [6] | 3.59 | 4.54 | 9.03 | 18.14 | 32.70 | 36.11 | 37.61% | 47.88% | 56.27% | 64.90% | 70.90% | 82.07% | 87.60% | 89.49% |
| ORB [16] | 2.36 | 6.84 | 9.49 | 21.36 | 36.17 | 38.83 | 18.54% | 29.30% | 41.05% | 52.77% | 31.69% | 51.24% | 66.84% | 76.26% |
| MSER [9] | 2.67 | 8.97 | 8.77 | 20.07 | 37.58 | 38.58 | 53.47% | 61.64% | 68.97% | 73.01% | 68.66% | 83.34% | 87.77% | 90.70% |

Table 2: Feature tracking error (Eqn. 5), illumination repeatability error (%) (Eqn. 6) and noise sensitivity error (%) (Eqn. 6)

$$error = \frac{\sum_{i=1}^n \sum_{j=1}^k |v - s_{ij}|}{kn} \quad (5)$$

where v is the predefined vibration amplitude, s is the measured displacement of each feature, k denotes the number of detected features and n denotes the total number of frames in the test image sequence. Table 2 shows the results, based on Eqn. 5, and shows that most prior approaches demonstrated similar performance under vibration with both DeGraF- α and DeGraF- β offering superior performance.

Illumination invariance is evaluated using image sequences with variable brightness settings. Given a set of features detected in an original image, the detection error is derived by measuring the repeatability of features in images adjusted for a 25%, 50%, 75% and 100% higher illumination level. The resulting repeatability error can be expressed as follows:

$$error = \frac{\|A \cup B\| - \|A \cap B\|}{\|A \cup B\|} \quad (6)$$

based on the difference in magnitude between the union of all the spatial features detected before (A) and after (B) the change in image conditions and their spatial intersection (based on position). Table 2 shows the mean repeatability error caused by illumination variance normalized as a percentage of all features. DeGraF- β outperforms all other methodologies with a consistently low error, while DeGraF- α shows comparable performance at higher brightness levels. Notably, ORB has the third lowest error rate and is the only other approach apart from DeGraF that is based on intensity centroids, albeit at low density.

Our noise stability evaluation adds noise to each image with $n\%$ of pixels having Gaussian distribution ($\sigma = 1$) noise added and feature repeatability measured as per Eqn. 6. Table 2 clearly illustrates that DeGraF- β outperforms all other approaches by a significant margin. Comparing DeGraF- β to the second best approach (ORB) shows a performance gap between 28% and 62% when image noise incidence increases from 5% and 20% respectively. Again, the fact that ORB and DeGraF- α are second/third in the performance ranking is notable, since they are the only other approaches that use intensity-weighted centroids for feature extraction.

Finally, we measure rotation invariance using artificial image rotation for $\pm\theta \in \{-3...3\}^\circ$ (corresponding to the typical

| | image rotation by $\pm\theta \in \{-3...3\}$ degrees | | | | | |
|----------------------------------|--|---------------|---------------|---------------|---------------|---------------|
| | -3 | -2 | -1 | 1 | 2 | 3 |
| AGAST [13] | 54.87% | 56.81% | 51.87% | 51.19% | 54.66% | 55.78% |
| CenSurE [12] | 65.91% | 68.40% | 57.35% | 58.11% | 67.05% | 64.53% |
| DeGraF- α | 65.24% | 65.68% | 63.21% | 63.39% | 65.60% | 64.84% |
| DeGraF-β | 41.81% | 42.30% | 41.49% | 41.11% | 42.21% | 41.42% |
| FAST [3] | 58.13% | 58.26% | 53.11% | 53.27% | 57.80% | 57.28% |
| GFTT [5] | 49.51% | 51.17% | 42.54% | 42.82% | 49.82% | 47.92% |
| SIFT [4] | 50.96% | 51.93% | 45.25% | 45.35% | 51.19% | 49.21% |
| SURF [6] | 66.19% | 65.69% | 56.25% | 56.16% | 65.14% | 64.08% |
| ORB [16] | 43.42% | 44.45% | 38.00% | 38.24% | 43.99% | 41.98% |
| MSER [9] | 66.51% | 66.60% | 58.05% | 57.84% | 64.57% | 64.07% |

Table 3: Rotation repeatability error (%) (see Eqn. 6)

characteristics of road vehicles) with a corresponding feature repeatability error measures as per Eqn. 6. From the results of Table 3, we observe ORB and DeGraF- β have the highest rotation invariance, albeit with significant error of around 41%. It is notable, that poorly performing SURF and DeGraF- α use rectangular image areas to derive the properties of each feature, thus inherently limiting their rotational invariance.

Overall we see strong relative performance of both the DeGraF- α / DeGraF- β methods, strongly correlated by the performance of the ORB intensity-weighted centroid approach, with the DeGraF- β outperforming other state of the art methods based on the standard feature stability and repeatability tests [11] against ground truth. Qualitative evaluation is shown in Figures 1 and 4.

4. CONCLUSION

This work introduces DeGraF (Dense Gradient-based Features) approach for the efficient computation of dense scene features based localized intensity centroids. This approach is shown to significantly outperform a range of state of the art approaches in the field ([3, 4, 9, 12, 6, 5, 15, 16]) in terms of feature density, computational efficiency (translating to high frame-rates) and overall feature stability under variation in terms of frame-to-frame tracking, orientation, noise and changes in illumination. Future work will consider further aspects of recent advances analogous sparse features [17, 18].

[This work is supported by ZF TRW Conekt and EPSRC]

5. REFERENCES

- [1] C. Harris and M. Stephens, "A combined edge and corner detector," *Proc 4th Alvey Vision Conf.*, 1988.
- [2] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," *Order A Journal On The Theory Of Ordered Sets And Its Applications*, vol. 7597, no. 7597, p. 22, 1991.
- [3] J. Shi and C. Tomasi, "Good features to track," *Proc. Conf. on Computer Vision and Pattern Recognition*, vol. 94, no. June, pp. 593–600, 1994.
- [4] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [5] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection.," *IEEE Trans. on Patt. Anal. Mach. Intel.*, vol. 32, no. 1, pp. 105–19, 2010.
- [6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [7] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. European Conf. Computer Vision*, pp. 128–142, 2002.
- [8] K. Mikolajczyk, "Scale & Affine Invariant Interest Point Detectors," *Int. Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [9] J. Matas, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
- [10] T. Kadir, A. Zisserman, and M. Brady, "An Affine Invariant Salient Region Detector," in *European Conf. Computer Vision*, pp. 228–241, 2004.
- [11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A Comparison of Affine Region Detectors," *Int. Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.
- [12] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *Proc. European Conf. Computer Vision*, pp. 102–115, Springer, 2008.
- [13] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and Generic Corner Detection Based on the Accelerated Segment Test," in *Proc. European Conf. Computer Vision*, pp. 183–196, Springer, 2010.
- [14] J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.
- [15] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 34, no. 7, pp. 1281–1298, 2012.
- [16] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *Int. Conf. on Computer Vision*, pp. 2564–2571, 2011.
- [17] A. Alahi, R. Ortiz, and P. Vanderghenst, "Freak: Fast retina keypoint," in *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 510–517, IEEE, 2012.
- [18] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Computer Vision*, pp. 2548–2555, IEEE, 2011.
- [19] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, IEEE, 2012.
- [20] L. Zhou, Z. Zhou, and D. Hu, "Scene classification using a multi-resolution bag-of-features model," *Pattern Recognition*, vol. 46, pp. 424–433, jan 2013.
- [21] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," in *Intelligent Vehicles Symposium*, pp. 486–492, 2010.
- [22] R. A. Newcombe and A. J. Davison, "Live dense reconstruction with a single moving camera," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1498–1505, IEEE, jun 2010.
- [23] A. Wedel, C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers, *Efficient dense scene flow from sparse or dense stereo data*. Springer, 2008.
- [24] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.
- [25] F. Mroz and T. Breckon, "An empirical comparison of real-time dense stereo approaches for use in the automotive environment," *EURASIP Journal on Image and Video Processing*, vol. 2012, no. 13, pp. 1–19, 2012.
- [26] O. Hamilton, T. Breckon, X. Bai, and S. Kamata, "A foreground object based quantitative assessment of dense stereo approaches for use in automotive environments," in *Proc. International Conference on Image Processing*, pp. pp. 418–422, IEEE, September 2013.
- [27] P. L. Rosin, "Measuring corner properties," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 291–307, 1999.
- [28] F. Ahmad, M. Gilbert, S. Myers, J. Pacheco, R. Castellane, and E. Miller, "Lagrange Gradient Mask for Optical Image Processing," *The Open Optics Journal*, vol. 1, no. 1, pp. 4–7, 2007.
- [29] S. Avidan, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients," in *Proc. Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 1491–1498, IEEE, 2006.
- [30] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, no. 3, pp. 886–893, 2004.
- [31] I. Katramados and T. Breckon, "Real-time visual saliency by division of gaussians," in *Proc. Int. Conf. on Image Processing*, pp. 1741–1744, IEEE, 2011.
- [32] T. Vaudrey, C. Rabe, R. Klette, and J. Milburn, "Differences between stereo and motion behaviour on synthetic and real-world stereo sequences," *Int. Conf. Image and Vision Computing New Zealand*, pp. 1–6, 2008.
- [33] J. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 1, no. 2, pp. 1–9, 2001.